



Revista alergia México

ISSN: 0002-5151

ISSN: 2448-9190

Colegio Mexicano de Inmunología Clínica y Alergia, A.C.

Rendón-Macías, Mario Enrique; Villasís-Keever, Miguel Ángel  
Fases para determinar la utilidad clínica de las pruebas diagnósticas  
Revista alergia México, vol. 67, núm. 3, 2020, Julio-Septiembre, pp. 279-285  
Colegio Mexicano de Inmunología Clínica y Alergia, A.C.

DOI: <https://doi.org/10.29262/ram.v67i3.831>

Disponible en: <https://www.redalyc.org/articulo.oa?id=486771931007>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en [redalyc.org](https://www.redalyc.org)



Sistema de Información Científica Redalyc  
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso  
abierto

## Phases to determine the clinical utility of diagnostic tests

### Fases para determinar la utilidad clínica de las pruebas diagnósticas

Mario Enrique Rendón-Macías,<sup>1</sup> Miguel Ángel Villasís-Keever<sup>2</sup>

#### Abstract

In general, diagnostic tests are the necessary tools to determine the presence or absence of illness, but they can be useful also for excluding other differential conditions, assessing severity, establishing specific treatments, and estimating possible prognosis results. In order to determine how much a new diagnostic test can contribute in the clinical setting, studies with different research designs and populations are required. Since it might be complex to understand the development process of a diagnostic test based on its initial stages, this article synthesizes and gives examples of the four phases that have been proposed to define this process. The goals of these four phases are: phase I, establishing reference values; phase II, analyzing the validity of the test; phase III, the impact of its incorporation in diagnostic-therapeutic plans; and phase IV, a long-term assessment after incorporating the diagnostic test.

**Key words:** Diagnostic tests; Development phases; Validation

**Este artículo debe citarse como:** Rendón-Macías ME, Villasís-Keever MÁ. Fases para determinar la utilidad clínica de las pruebas diagnósticas. Rev Alerg Mex. 2020;67(3):279-285

#### ORCID

Mario Enrique Rendón-Macías, 0000-0001-7310-6656; Miguel Ángel Villasís-Keever, 0000-0002-8566-0811

<sup>1</sup>Universidad Panamericana, Escuela de Medicina, Ciudad de México, México

<sup>2</sup>Instituto Mexicano del Seguro Social, Centro Médico Nacional Siglo XXI, Unidad de Investigación en Análisis y Síntesis de la Evidencia, Ciudad de México, México

Correspondencia: Miguel Ángel Villasís-Keever.  
miguel.villasis@gmail.com

Recibido: 2020-06-24

Aceptado: 2020-07-01

DOI: 10.29262/ram.v67i3.831



## Resumen

Generalmente, las pruebas de diagnóstico son las herramientas necesarias para determinar la presencia o no de enfermedad, pero también pueden ser útiles para la exclusión de otras entidades diferenciales, evaluar la gravedad, establecer tratamientos específicos y estimar posibles resultados pronósticos. Para determinar cuánto puede contribuir una nueva prueba de diagnóstico en el ámbito clínico, se requieren estudios con diseños de investigación y poblaciones diferentes. Dado que puede resultar complejo entender el desarrollo de una prueba diagnóstica a partir de sus fases iniciales, en este artículo se sintetizan y se proporcionan ejemplos de las cuatro fases que se han propuesto para definir este proceso. Las metas de estas etapas son las siguientes: fase I, establecimiento de valores de referencia; fase II, análisis de la validez de la prueba; fase III, impacto de su incorporación en el plan diagnóstico; fase IV, evaluación a largo plazo de la incorporación de la prueba diagnóstica.

**Palabras clave:** Pruebas diagnósticas; Fases de desarrollo; Validación

## Abreviaturas y siglas

ABC, área bajo la curva

DE, desviación estándar

FEF<sub>25-75</sub>, flujo espiratorio medio

FeNO, fracción exhalada de óxido nítrico

IC, intervalo de confianza

NNDA, número necesario para un diagnóstico adecuado

NNED, número necesario para un error diagnóstico

ROC, receptor operante

## Introducción

Las pruebas de diagnóstico permiten acercarnos a una valoración más adecuada de la condición clínica de un paciente para, generalmente, determinar si tiene o no una enfermedad y, por lo tanto, facilitan la toma de decisiones en beneficio del paciente. En el ámbito clínico, la decisión de usar o no una prueba o grupo de pruebas de diagnóstico se debe basar en evidencia científica, la cual es alcanzada mediante diferentes estudios dirigidos a contestar preguntas específicas sobre la *validez* de una nueva prueba diagnóstica. Para determinar qué tan útil y válida es una prueba diagnóstica se requiere un proceso, es decir, la realización de diversos estudios de investigación, con preguntas específicas, donde se incluyen diferentes poblaciones, diseños y análisis estadísticos.

En ese sentido, es necesario comentar que los resultados de una prueba diagnóstica no solo sirven para conocer la sensibilidad, especificidad y valores predictivos para confirmar o descartar una enfermedad, sino que tienen otros propósitos, como el escrutinio o tamizaje, la evaluación del comportamiento de una enfermedad (es decir, el monitoreo para analizar el control adecuado, la remisión o la recaída), o bien, para estimar el pronóstico de las

enfermedades. Debido a lo complejo que puede ser este proceso, aproximadamente durante 20 años se consideró apropiado proponer fases para el desarrollo de una nueva prueba diagnóstica a fin de hacerlo más comprensible,<sup>1,2,3</sup> en forma semejante a lo que ocurre cuando se está evaluando un fármaco nuevo.<sup>4</sup>

En general se ha establecido que cuatro fases permiten llevar un orden en el conocimiento de la nueva prueba diagnóstica, a partir de ahora referida como “prueba índice”. En las fases iniciales se busca establecer la confiabilidad y validez de la prueba índice para identificar la condición clínica de un paciente, es decir, sano o enfermo; mientras que en las fases posteriores se establece la repercusión en la salud una vez que la prueba índice se aplica en condiciones reales.<sup>1,3,5</sup> Hay ciertas variaciones a la propuesta original de esas cuatro fases, las cuales se presentan en el cuadro 1.

## Fases de estudio en una prueba diagnóstica

**Fase I. Información sobre los valores normales y su potencial uso diagnóstico**

Los primeros estudios para proponer una nueva prueba índice contestarán a la pregunta ¿cuán fre-

Cuadro 1. Características clínicas de pacientes con síndrome de alergia oral			
Fases	Objetivo	Diseño	Características participantes
Fase I	Establecer los valores normales	Transversal descriptivo	Sujetos sanos
Fase II	Ila Comparar valores entre dos grupos	Casos y controles	Sanos y enfermos
	Ilb Determinar la sensibilidad y especificidad de la prueba índice	Casos y controles	Sanos y enfermos con una condición específica (situación clínica ideal)
	Ilc Calcular valores predictivos	Transversal comparativo	Sujetos con sospecha de la enfermedad (situación clínica real)
Fase III	Evaluación de las consecuencias clínicas al usar la prueba índice nueva	Ensayos clínicos controlados	Sujetos que reciben o no reciben el resultado de la prueba índice, de acuerdo como fueron identificados por la prueba (sanos o enfermos)
Fase IV	Estimación de las consecuencias a largo plazo del uso de la prueba índice nueva	Estudios de cohorte	Sanos y enfermos en quienes se utiliza la prueba diagnóstica en la práctica habitual

cuenta es el marcador clínico o molecular en sujetos sin la enfermedad de interés? De la misma forma es necesario plantear ¿por qué esta prueba índice se modificaría ante una enfermedad específica? Con las respuestas a estas dos preguntas se conocerá si una nueva prueba índice podrá tener impacto en el proceso de diagnóstico de una enfermedad específica.

Para resolver la primera pregunta es necesario establecer los valores de “normalidad” o de referencia en condiciones sin la enfermedad en cuestión. Los estudios encaminados a contestar esta primera pregunta deben ser observacionales dirigidos a poblaciones de sujetos sanos y, en general, transversales, aunque los estudios de cohorte son una opción cuando se desea evaluar cambios en un tiempo. De preferencia deben incluir grupos grandes para lograr una mejor representación de la población general. En particular, si la prueba índice se plantea en un futuro para ser aplicada en enfermedades propias de algunos grupos de edad, estos individuos deben estar representados en una gran cantidad (recomendamos muestras de al menos 500 personas).

Desde esta primera fase es importante que los autores informen con detalles sobre las condiciones de la recolección de los datos, la o las técnicas utilizadas, los equipos usados y los criterios de análisis de los resultados. Asimismo, es necesario que las mediciones realizadas en cada individuo con la prueba

índice sean reproducibles, es decir, el resultado de cada individuo no cambie al repetir el estudio (repetibilidad, fiabilidad o consistencia). Este análisis puede ser reportado con medidas estadísticas como el coeficiente kappa (cuando se trata de mediciones cualitativas), el coeficiente de correlación intraclase o el coeficiente de variación y gráficas de Bland y Altman (para medidas cuantitativas).

El análisis de los datos debe informar sobre la distribución de estos. En las variables cualitativas se deberá indicar la frecuencia relativa (porcentaje) de sujetos que cumplieron con el criterio aceptado como normal; este es un criterio clínico y no estadístico. En las variables cuantitativas se mostrará la distribución de los valores para establecer si cumplen criterios estadísticos de asimetría (“normal”) o asimétrica. En condiciones de simetría, los valores se presentarán con la media y su desviación estándar (DE).

Para cubrir la probabilidad de variación del dato (prueba índice) en 95 % de la población sana, los valores de referencia se establecerán con un valor límite inferior, definido como el valor encontrado a  $-1.96$  DE de la media y con un valor límite superior a  $+1.96$  DE. En caso de una distribución asimétrica, los valores normales se resumirán en medianas y percentiles; el valor de referencia límite menor será el correspondiente al percentil 2.5 (2.5 %) y el mayor, al percentil 97 (97 %), con lo cual se cubre 95 % de la variación en la población.<sup>6</sup>

La segunda pregunta de la fase I suele contestarse con estudios básicos o experiencias clínicas previas. En los estudios básicos se revisan los mecanismos etiopatogénicos y fisiopatológicos de las enfermedades y, con ello, las moléculas con potencialidad de marcadores diagnósticos. En experiencias con casos clínicos informado, se determinan los datos clínicos susceptibles de ser medidos y registrados con métodos clínicos o de imagen para su postulación como herramientas diagnósticas.

### Ejemplo

En una evaluación de diferentes centros, Buchvald *et al.*<sup>7</sup> establecieron los valores normales de la fracción exhalada de óxido nítrico (FeNO). Para ello, lo midieron con un analizador de quimioluminiscencia en 522 sujetos menores de 17 años. Para establecer los valores normales calcularon la media geométrica y el percentil 95 % o límite superior aceptable, el cual se fijó en < 15 o 25 ppb conforme la edad.

### Fase II. Evaluación de la validez diagnóstica de la prueba índice

El objetivo fundamental de una prueba diagnóstica es su capacidad para discriminar entre las personas que tienen una condición, habitualmente una enfermedad, de quienes no la presentan. Una prueba índice se dice que es totalmente válida si su resultado siempre informa correctamente la presencia o ausencia de esa condición. Para probar esta situación es necesario contrastar la nueva prueba índice con un estándar infalible (también llamado “estándar de oro”) que establece la condición real (con o sin la enfermedad) de una persona.

Las preguntas en esta fase son las siguientes: ¿los resultados en la prueba índice difieren entre las personas con (enfermos) y sin esta condición (sanos)?, ¿las personas con la condición establecida (enfermos) difieren de quienes tienen enfermedades parecidas (diferenciales)?, ¿las personas con una condición (enfermedad) de presentación grave tienen resultados más “anormales” que quienes tienen una condición leve?, ¿cuánto se modifica la capacidad discriminativa de la prueba índice en personas con otras condiciones (enfermedades) presentes?, ¿cómo se comporta la prueba índice ante complicaciones en una misma enfermedad? y ¿la prueba índice tendrá mejor desempeño en conjunto con otras pruebas?

Para contestar estas preguntas se utilizan diseños transversales comparativos (cuadro 1), los cuales deben asegurar el cumplimiento de cuatro aspectos:

1. Incluir individuos con la enfermedad y controles (ver más adelante) en quienes se determine su condición clínica (aplicación del estándar de oro) y la prueba índice por evaluar. Se ha propuesto una subfase IIa en la que el grupo con la enfermedad sea integrado por individuos con la enfermedad, tanto en condiciones leves como severas de evolución, contrastando con un grupo control de sujetos sanos; así como una subfase IIb, en la que el grupo control lo integrarían sujetos con otras comorbilidades e incorporando sujetos con enfermedades diferenciales a la de interés. Finalmente, una IIc, que incluiría sujetos con la enfermedad o sospecha de la misma y con enfermedades diferenciales; aquí se incluyen la aplicación de otras pruebas de diagnóstico.
2. Es necesario que los resultados de la prueba índice sean determinados ignorando la condición del paciente (resultado del estándar de oro). Se debe informar sobre la reproducibilidad de la prueba índice, así como la consistencia intra e interevaluador(es) cuando el resultado requiera una interpretación; e informar sobre las estrategias de cegamiento usadas para asegurar la independencia de los resultados.
3. Las evaluaciones o mediciones de ambas condiciones deben ser concurrentes, es decir, en un tiempo en el cual las condiciones del paciente no hayan cambiado entre una y otra medición.
4. Asegurar un número suficiente de participantes en todas las categorías diagnósticas para lograr emitir estimados confiables de los índices de validez.

En esta fase se reporta los índices de validez de una nueva prueba diagnóstica. Estos dependen de la forma en la cual se mide el resultado de la prueba índice. Si el resultado se emite como una variable dicotómica (positiva contra negativa), los índices se pueden calcular con una tabla de  $2 \times 2$  (enfermos contra sanos, y positivos contra negativos). A través de esta tabla se establece la *sensibilidad* o validez de la prueba en los enfermos, la *especificidad* o validez de la prueba en los controles, así como la *precisión* o validez global de la prueba (capacidad para identi-

ficar correctamente tanto a los enfermos como a los sanos). En general, valores  $\geq 0.85$  o 85 % de cada uno traducen muy buena validez.<sup>8</sup> Otras formas de informar sobre la validez en esta condición son los cocientes (o razones) de verosimilitud o probabilidad. Estos indican la probabilidad de presentar una enfermedad contra no tenerla cuando una prueba es positiva y cuando es negativa; valores superiores a 10 o 1/10 indican alta validez de la prueba índice. Finalmente, en años recientes se ha propuesto la razón de oportunidades (razón de momios, *odds ratio*) diagnóstica como una medida de asociación entre la positividad con una prueba y el riesgo de estar enfermo; valores  $> 3$  son altamente sugestivos de la enfermedad. Para un resultado en una escala ordinal, dada la posibilidad de más de una respuesta (por ejemplo, enfermedad confirmada, resultado dudoso y resultado negativo), se recomienda el uso de los cocientes de probabilidad.

En un resultado cuantitativo son requeridos dos análisis para evaluar la validez de una nueva prueba diagnóstica. La primera es la construcción de una curva del receptor operante (curva ROC) y la determinación de su área bajo la curva (ABC). Si esta área es de 100 % (1.0), la prueba es totalmente válida. El segundo análisis consiste en aprovechar la construcción de la curva ROC y determinar el nivel de corte (valor para emitir un resultado positivo contra uno negativo) que determina el punto donde la prueba muestra el mejor desempeño diagnóstico, es decir, su mayor sensibilidad y especificidad.

Otros análisis que se pueden informar en esta fase son los relacionados con el posible impacto para el diagnóstico, tales como los valores predictivos positivo (probabilidad de estar enfermo dado que se tiene un resultado positivo) y negativo (probabilidad de no tener la condición o enfermedad dado que se tiene un resultado negativo), la probabilidad posprueba con el uso de los cocientes de verosimilitud o probabilidad, el número necesario para el diagnóstico y el error diagnóstico y, finalmente, el índice de utilidad clínica.

### Ejemplos

Para determinar si el FeNO en condiciones basales y en respuesta a broncodilatador pudiera ser útil como una prueba diagnóstica para identificar niños preescolares con probable asma, Malmberg *et al.*<sup>7,9</sup> estudiaron a 96 niños con síntomas de asma o histo-

ria clínica de la misma comparados con 62 controles sanos, pareados por edad. En un estudio transversal compararon los niveles entre los grupos, determinaron la posible utilidad diagnóstica a través de la construcción de una curva ROC y la determinación de un punto de corte para la mejor sensibilidad y especificidad de la prueba. Primero comprobaron que los grupos tenían valores de distribución diferente: promedio  $22.1 \pm 3.4$  ppb *versus*  $5.3 \pm 0.4$  ppb. Conforme el análisis de la curva ROC con un criterio de 1.5 DE sobre el valor predicho, la FeNO tenía una validez adecuada, sensibilidad de 86 % y una especificidad de 92 %.

Por su parte, Eom *et al.*<sup>10</sup> evaluaron si la combinación única de los índices espirométricos con la medición de la FeNO en la primera consulta sería suficiente para realizar el diagnóstico de asma en niños. Estudiaron a 275 niños de ocho a 16 años de edad, de los cuales a 191 se les diagnosticó asma después de una evaluación completa (antecedentes, examen clínico, pruebas de alergia, medición de FeNO, espirometría con prueba de reversibilidad a broncodilatador). Determinaron el desempeño a través del análisis del ABC. La FeNO mostró un área de 0.79 (0.74-0.85), pero mejoró su desempeño con la adición del flujo espiratorio medio ( $FEF_{25-75}$ ) a 0.90 ( $FEF_{25-75}$  95 % = 0.86-0.93).

### Fase III. Evaluación de las consecuencias clínicas de introducir una nueva prueba diagnóstica

El que una nueva prueba índice sea válida no implica en sí mismo que su uso rutinario pudiera cambiar el pronóstico de forma favorable en los pacientes sometidos a ella. En ocasiones, la decisión apresurada pudiera ser contraproducente en la atención integral de los pacientes. En esta fase, las investigaciones sobre una nueva prueba diagnóstica pretenden contestar las siguientes preguntas: ¿los pacientes evaluados con la nueva prueba muestran mejores resultados clínicos en comparación con los evaluados con la prueba estándar o, incluso, si no se les hace ninguna prueba?, ¿se observan diferencias en el tratamiento y la consecuencia del mismo cuando un paciente sale positivo a la nueva prueba contra otro diagnosticado con la prueba estándar o habitual?, ¿influye en el tratamiento y sus consecuencias, el ser estudiado primero con la prueba diagnóstica nueva que realizarla después de la habitual? En la respuesta a estas inte-



rogantes no solo debe considerarse el efecto positivo sino también las posibles consecuencias nocivas.

Para contestar a estas preguntas, en la fase III se recomienda realizar estudios de ensayos clínicos aleatorizados. En estas investigaciones se comparan participantes sometidos a la prueba índice nueva con otro grupo similar, pero en quienes se usa la prueba diagnóstica habitual; la asignación de un grupo a otro es aleatoria. Asimismo, la evaluación de las consecuencias en el tratamiento y el pronóstico deben realizarse de forma objetiva y de preferencia por personal sin conocimiento de la estrategia diagnóstica realizada (cegado). Es recomendable un estudio con el suficiente tamaño de muestra para alcanzar resultados consistentes y capaces de mostrar efectos secundarios poco frecuentes. La forma de evaluar estos estudios puede ser:

- Incluyendo pacientes con sospecha de la enfermedad en cuestión que requieren la prueba diagnóstica-terapéutica. De forma aleatoria se decide cuál prueba diagnóstica realizar y solo se proporcionará tratamiento si la prueba diagnóstica aplicada resulta positiva. Los resultados son presentados con estimadores de asociación (riesgo relativo, razón de momios o cociente de riesgos) o de impacto NNDA (número necesario para un diagnóstico adecuado), NNED (número necesario para un error diagnóstico).
- Incluyendo pacientes con sospecha de la enfermedad en cuestión y seguimiento también a los negativos a la prueba diagnóstica-terapéutica. En este diseño se observa la consecuencia de un resultado negativo de la prueba diagnóstica realizada (nueva o habitual); los pacientes con estos resultados son seguidos para observar si hay recaída o recrudescimiento de la enfermedad a consecuencia de una condición de falso negativo.
- Incluir solo pacientes con la enfermedad confirmada para modificación del tratamiento. Este tipo de diseños se realiza con el objetivo de analizar si la nueva prueba modificaría el control de una enfermedad bajo un tratamiento. Los pacientes en tratamiento se someten a la nueva prueba índice o la habitual y, si el resultado es positivo, el tratamiento se cambia; de lo contrario, continúa igual. En ambos grupos se determina las tasas de curación o control, así como la tasa de complicaciones.

En esta fase se pueden realizar estudios bajo una perspectiva de análisis económico. Estos pueden ser de costo-efectividad o impacto económico de la nueva prueba en los resultados clínicos; de costo-utilidad o impacto evaluado sobre la calidad de vida; o de costo-beneficio o impacto, para evaluar, por ejemplo, el dinero ahorrado o gastado.

### Ejemplo

Boer *et al.*<sup>11</sup> analizaron la evolución de los pacientes de consulta de primer contacto tratados por asma con dos esquemas de inhibidores de la enzima convertidora de angiotensina conforme sus niveles de FeNO. En ambos grupos se revisó control del asma, calidad de vida relacionada con asma, medicamentos y costo de las estrategias. A los 12 meses de tratamiento, en los pacientes con niveles < 25 ppb de FeNO basal se encontró una reducción significativa en la dosis de beclometasona, en promedio de 223 µg (IC 95 % = 6-439). Además, en este grupo el costo del tratamiento se redujo 159 dólares americanos (IC 95 % = 33-285). En niveles superiores de FeNO no hubo diferencias. Tampoco se observó diferencias en la calidad de vida en ambos grupos.

Fase IV. Consecuencias a largo plazo de la introducción de la nueva prueba en la práctica clínica. La continuación en la adopción de una nueva prueba diagnóstica es determinada después de ser incluida dentro de las estrategias diagnósticas con sus respectivas consecuencias terapéuticas.

En esta fase, la forma de evaluar la repercusión de la incorporación de la nueva prueba diagnóstica en la atención habitual de las personas se puede hacer a través de estudios de cohortes o ensayos clínicos. En estos estudios, la comparación se hace en estudios ecológicos o de campo en los cuales se analizan las repercusiones en salud pública de la incorporación de la nueva prueba con su respectivo impacto sobre el tratamiento. Los estudios son menos controlados que en la fase III y pueden analizar tanto el beneficio como los daños con baja incidencia o sucedidos en tiempos más prolongados.

### Ejemplo

Posterior a los diferentes estudios mencionados y ya usado en la práctica clínica, un grupo de expertos (neumólogos, expertos en fisiología de la respiración y alergólogos) analizaron el uso de FeNO como biomarcador en el diagnóstico y tratamiento del asma.

Mediante método Delphi y discusión grupal llegaron a las siguientes conclusiones:<sup>12</sup>

[...] la medición de la FeNO sirve para el diagnóstico de asma en clínicas especializadas, tanto en niños como adultos, así como para determinar el nivel de tratamiento con corticosteroides. En asma grave, recomendamos la FeNO para la endotipificación, detectar la mala adherencia terapéutica, el subtratamiento y el riesgo de crisis. Sugerimos su uso para determinar el nivel de tratamiento con corticosteroides e identificar pacientes con riesgo de tener una pérdida de la función pulmonar. También la recomendamos en el adulto para mejorar la elección de medicamentos biológicos y, en este contexto, solo la sugerimos en casos selectos en niños.

## Conclusiones

La decisión de incorporar una nueva prueba diagnóstica al arsenal clínico en cualquier enfermedad implica su validación y evaluación a través de varios estudios, tanto para determinar su desempeño (sensibilidad y especificidad), así como para establecer su repercusión en la toma de decisiones clínicas, al ser parte de las pruebas diagnósticas habituales. Para lograr una evaluación integral de estos aspectos de manera más o menos ordenada se han propuesto cuatro fases: determinación de los valores de referencia, comprobación de la validez de la prueba, análisis del impacto de su inclusión al plan diagnóstico y evaluación de la repercusión a largo plazo de su incorporación.

---

## Referencias

1. Bolboacă SD. Medical diagnostic tests: A review of test anatomy, phases, and statistical treatment of data. *Comput Math Methods Med.* 2019;2019:1891569. DOI: 10.1155/2019/1891569
2. Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ.* 2002;324(7336):539-541. DOI: 10.1136/bmj.324.7336.539
3. Colli A, Fraquelli M, Casazza G, Conte D, Nikolova D, Duca P, et al. The architecture of diagnostic research: from bench to bedside--research guidelines using liver stiffness as an example. *Hepatology.* 2014;60(1):408-418. DOI: 10.1002/hep.26948
4. Zurita-Cruz JN, Barbosa-Cortés L, Villasís-Keever M. De la investigación a la práctica: fases clínicas para el desarrollo de fármacos. *Rev Alerg Mex.* 2019;66(2):246-253. DOI: 10.29262/ram.v66i2.625
5. Gluud C, Gluud LL. Evidence based diagnostics. *BMJ.* 2005;330(7493):724-726. DOI: 10.1136/bmj.330.7493.724
6. Rendón-Macías ME, Villasís-Keever MÁ, Miranda-Navales MG. Estadística descriptiva. *Rev Alerg Mex.* 2016;63(4):397-407. DOI: 10.29262/ram.v63i4.230
7. Buchvald F, Baraldi E, Carraro S, Gaston B, de Jongste J, Pijnenburg MWH, et al. Measurements of exhaled nitric oxide in healthy subjects age 4 to 17 years. *J Allergy Clin Immunol.* 2005;115(6):1130-1136. DOI: 10.1016/j.jaci.2005.03.020
8. Rendón-Macías ME, Valenzuela M, Villasís-Keever M. Sesgos en los estudios de pruebas de diagnóstico: implicación en la estimación de la sensibilidad y especificidad. *Rev Alerg Mex.* 2020;67(2):165-173. DOI: 10.29262/ram.v67i2.771
9. Malmberg LP, Pelkonen AS, Haahtela T, Turpeinen M. Exhaled nitric oxide rather than lung function distinguishes preschool children with probable asthma. *Thorax.* 2003;58(6):494-499. DOI: 10.1136/thorax.58.6.494
10. Eom SY, Lee JK, Lee YJ, Hahn YS. Combining spirometry and fractional exhaled nitric oxide improves diagnostic accuracy for childhood asthma. *Clin Respir J.* 2020;14(1):21-28. DOI: 10.1111/crj.13095
11. Boer S, Honkoop PJ, Loijmans RJB, Snoeck-Stroband JB, Assendelft WJJ, Schermer TRJ, et al. Personalised exhaled nitric oxide fraction (FeNO)-driven asthma management in primary care: a subgroup analysis of the ACCURATE trial. *ERJ Open Res.* 2020;6(3):00351-2019. DOI: 10.1183/23120541.00351-2019
12. Larenas-Linnemann D, Gochicoa-Rangel L, Macías-Weinmann A, Soto-Ramos M, Luna-Pech JA, Elizondo-Ríos A, et al. Consenso mexicano en relación con la fracción exhalada de óxido nítrico (FeNO) en asma 2020. *Rev Alerg Mex.* 2020;67(Supl 2):S1-S25. DOI: 10.29262/ram.v67i0.760.